2K

УДК 004.942

А.С. Коряк
Харьковский национальный университет радиоэлектроники, г. Харьков, Украина
swell.kture.kharkov.ua

Информационно-аналитическая система поддержки банковского кредитования, основанная на плагинах

В статье приведено описание информационно-аналитической системы поддержки банковского кредитования, основанной на плагинах. Для описанной системы разработаны плагины, содержащие алгоритмы обучения, работы и визуализации нейронных сетей, предназначенных для построения моделей кластеризации, классификации и прогнозирования.

Введение

В последние десятилетия в мире бурно развивается банковская структура, стремительно увеличивается количество клиентов и предлагаемый банком спектр услуг, увеличивается уровень конкуренции между банками. Стратегическая стабильность банка на рынке возможна только при условии его конкурентоспособности и возможности адаптации к изменениям рыночной среды. Одним из наиболее действенных легальных методов укрепления конкурентных позиций в арсенале компаний является интеллектуальный анализ информации. На сегодняшний день актуальна задача разработки системы поддержки банковских решений, предоставляющей средства интеллектуального анализа для решения задач банковской деятельности.

Цель работы – одним из вариантов решения этой проблемы является создание модульной аналитической системы. Система должна поддерживать механизм подключения модулей, содержащих новые функции.

Рассмотрим построение системы поддержки процесса кредитования, основанной на модульности. Одним из возможных способов построения модульной и расширяемой архитектуры является использование плагинов. Плагин, или подключаемый модуль – независимо компилируемый программный модуль, динамически подключаемый к основной программе, предназначенный для расширения и/или использования её возможностей [1]. Плагины обычно выполняются в виде отдельных подключаемых библиотек. Например, плагином к графическому редактору может быть фильтр, который каким-либо образом изменяет изображение, палитру и прочее. Большой популярностью пользуются плагины для почтовых программ, а именно спам-фильтры, плагины для проверки писем антивирусом и др. Программы, широко известные своими интерфейсами для плагинов: Microsoft Outlook, Adobe Photoshop, Eclipse, Far Manager, Miranda IM, Firefox, Total Commander и др.

В основе архитектуры, основанной на плагинах, лежит ядро системы с набором интерфейсов, предоставляющих средства для доступа к объектной модели ядра системы. Ядро системы предоставляет общий набор функций, используемых всеми плагинами, например, функции доступа к базе данных, доступ к электронным письмам в почтовых программах, функции рисования в графических редакторах, доступ к объектной модели DOM в браузерах и html-редакторах.

Постановка задачи. В основе архитектуры системы поддержки кредитования должно лежать ядро, предоставляющее интерфейс с функциями для подключаемых модулей. Определим функции, предоставляемые ядром:

– обеспечение соединения с хранилищем данных (ХД) или другим источником данных;
– импорт из ХД данных по запросу пользователя;
– преобразование и нормализация данных для анализа;
– предоставление данных для плагинов в виде объектов-выборок, содержащих как сами данные, так и правила для преобразования и нормализации этих данных;
– загрузка плагинов по выбору пользователя;
– отображение различных пользовательских интерфейсов плагинов для различных категорий пользователей системы, например, аналитиков и кредитных инспекторов;
– сохранение и загрузка настроенных при помощи плагинов аналитических моделей.

Определим общую структуру и функции плагинов. Для разрабатываемой системы поддержки кредитования плагины содержат алгоритмы для интеллектуального анализа информации, например, такие как нейронные сети (НС) или деревья решений. Плагин должен реализовывать специальный интерфейс для взаимодействия с ядром. Реализация единого интерфейса всеми плагинами необходима для:

– идентификации. Ядро системы находит плагины по реализуемому интерфейсу;
– унификации плагинов. Благодаря единому интерфейсу для всех плагинов ядро может загрузить и использовать любой плагин независимо от того, какую функциональность этот плагин предоставляет.

Рассмотрим функции единого интерфейса, реализуемого всеми плагинами:

– предоставление интерфейсов для различных категорий пользователей;
– сериализация обученного и настроенного алгоритма в формат, пригодный для сохранения, например, в базе данных или XML-файле;
– загрузка и десериализация обученного и настроенного алгоритма из базы данных или файла.

Такой интерфейс не зависит от алгоритма интеллектуального анализа, который предоставляет плагин, что позволяет практически неограниченно расширять систему.

Плагины для интеллектуального анализа данных

На данный момент для описанной системы разработаны плагины, содержащие алгоритмы обучения, работы и визуализации нейронных сетей, предназначенных для построения моделей кластеризации, классификации и прогнозирования.

Выбор нейронных сетей в качестве плагинов для интеллектуального анализа обусловлен следующими факторами:

– нейронные сети имеют широкий круг применимости и позволяют решать сложные задачи прогноза, классификации или кластеризации, без которых не обходится ни один кредитный отдел и решение которых не могут осуществить статистические методы;
– нейронная сеть может на едином наборе входных данных решать несколько задач, имея несколько выходов, что даёт возможность кроме принятия решения о выдаче кредита, также рекомендовать подходящие данному потенциальному заемщику условия кредитования (например: сумму кредита и срок, на который он должен быть выдан) в соответствии с его материальным состоянием и другими характеристиками;
– одна нейронная сеть может быть применена для решения задач в различных предметных областях отдела кредитования банка, а также легко переобучена и адаптирована к другим условиям или другим задачам кредитования;

– нейронная сеть способна выделять нужные результаты из больших объёмов различной информации о заемщиках с большим числом характеристик и стабильно работать при наличии большого числа неинформативных данных;
– нейронная сеть может обучиться решению задачи, для которой у аналитика нет работающих с нужной точностью правил принятия решения, или аналитик отсутствует, что позволяет смягчить требования к сотрудникам отдела кредитования в банках.

Рассмотрим работу плагина для интеллектуального анализа на примере одной из нейронных сетей, входящих в состав описанной системы – самоорганизующихся карт Кохонена (Self-organizing map/SOM).

Плагин для SOM содержит алгоритм обучения нейронной сети Кохонена и методы визуализации карт, которые позволяют не только разделить выборку на кластеры, а затем классифицировать новые образы, но и визуально анализировать исследуемые данные с целью нахождения в них скрытых закономерностей.

Нейронная сеть Кохонена относится к самоорганизующимся сетям, которые при поступлении входных сигналов, в отличие от сетей, использующих обучение с учителем, не получают информацию о желаемом выходном сигнале. В связи с этим невозможно сформировать критерий настройки, основанный на рассогласовании реальных и требуемых выходных сигналов НС. Все предъявляемые входные сигналы из заданного обучающего множества самоорганизующаяся сеть в процессе обучения разделяет на классы, строя так называемые топологические карты [2].

В работе использован алгоритм обучения, описанный в [3]. В качестве функции соседства выбрана функция Гаусса, радиус в начале обучения – 3, норма обучения – 0,5.

Для визуализации карт Кохонена могут быть использованы 1-, 2- и 3-мерные пространства, но обычно практически ограничиваются отображением с помощью 2-мерных поверхностей, т.к. именно в таком виде человек воспринимает геометрические структуры наиболее естественно и отношения между объектами выглядят наиболее наглядно [4].

Под визуализацией данных картой Кохонена понимается такой способ представления многомерного распределения данных на двумерной плоскости, при котором качественно отражены основные закономерности, присущие исходному распределению – его кластерная структура, внутренние зависимости между признаками, информация о расположении данных в исходном пространстве и др.

Наиболее распространёнными 2-мерными способами визуализации карт Кохонена являются [5]: унифицированная матрица расстояний; проекция Саммона; матрица плотности попадания; матрица входов; матрица кластеров; матрица ошибок квантования и другие. Каждый из перечисленных способов визуализации карт Кохонена позволяет наблюдать те или иные свойства анализируемых данных.

Рассмотрим применение карт Кохонена для определения кластерной структуры исходных данных и выявления скрытых закономерностей в области кредитования на примере решения задачи кластеризации – разделения заемщиков на однородные группы со схожими характеристиками и их визуализации.

Набор исходных данных, использованный для кластеризации, состоит из 100 образов. Каждый образ представляет собой информацию о заемщиках и выданных кредитах и состоит из 30 характеристик, таких как: сумма кредита, срок кредита, цель кредитования, возраст, пол, образование, наличие квартиры, машины и др.

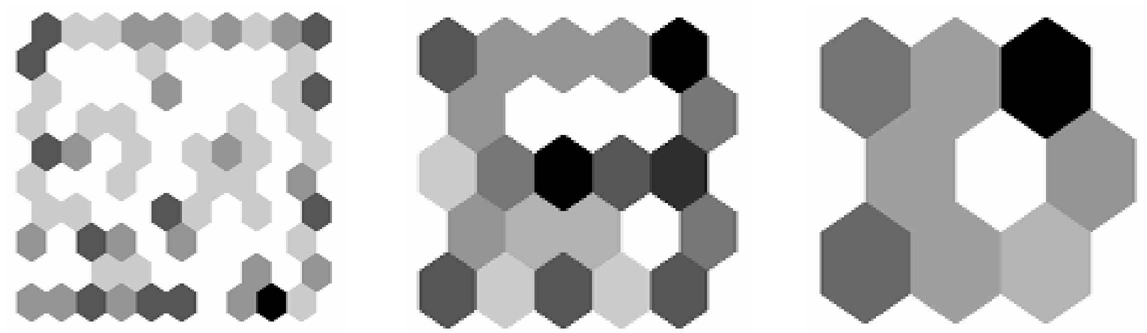
При реализации и интерпретации карт Кохонена важно понимать, что все карты – это разные раскраски одних и тех же нейронов. При этом каждый обучающий пример имеет одно и то же расположение на каждой из карт.

Также важным является выбор размера карты, т.е. количества нейронов. Некоторые источники [3] рекомендуют использовать максимально возможное количество нейронов в карте. В то же время проекция Саммона может быть применена непосредственно к набору данных, но является для вычисления очень интенсивной и актуальным становится применение нейронной сети Кохонена, т.к. SOM разбивает входные данные на малое количество векторов, что во многих случаях уменьшает нагрузку при вычислениях.

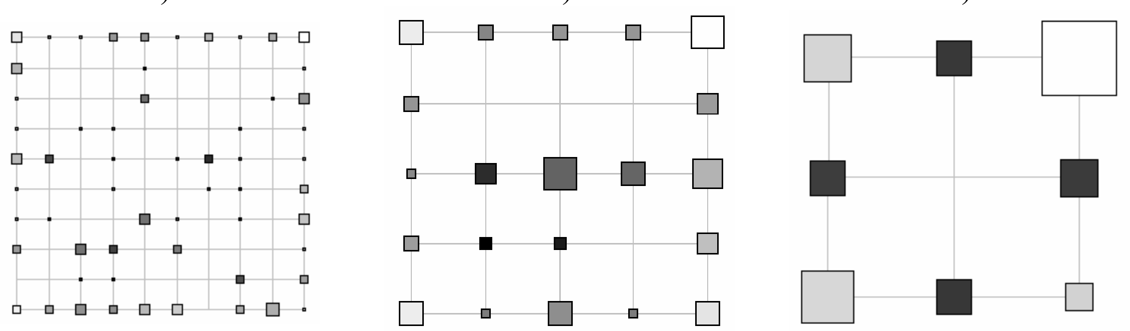
Кроме того, после обучения каждый входной пример попадает в «свой» нейрон. При этом в некоторые нейроны не попадет ни одного примера, а в некоторые попадет несколько примеров. Распределение обучающих примеров по нейронам очень показательно и отображается на карте частот, но более наглядной является диаграмма Хинтона. На рис. 1 представлены карты частот попадания и диаграммы Хинтона разного размера для описанного выше набора данных.

На картах частот попадания (рис. 1а, 1б, 1в) белым цветом показаны узлы решётки (нейроны), к которым не был отнесён ни один образ, чёрным – отнесено максимальное количество образов.

На диаграмме Хинтона на каждом узле сетки изображается квадрат, размер которого пропорционален числу точек, ближайших к данному узлу, а оттенок соответствует значению соответствующего отображаемого признака [4]. На диаграммах рис. 1г, 1д, 1е оттенок квадрата соответствует значению цвета на матрице расстояний для данного нейрона.



а) б) в)



г) д) е)

Рисунок 1 – Матрицы частот попадания и соответствующие им диаграммы Хинтона для карт размера: а) 10 × 10; б) 5 × 5; в) 3 × 3

Достоинством диаграммы Хинтона, кроме наглядности, является то, что оттенки узлов могут соответствовать необходимому признаку и диаграмма отображает сразу две карты, в данной ситуации – карту частот попадания и матрицу расстояний. Также квадраты можно заменить другими фигурами, в зависимости ещё от каких-то признаков.

Карта частот на рис. 1а и соответствующая диаграмма Хинтона (рис. 1г) размером 10 × 10 (100 нейронов) соответствует количеству образов исходной выборки. Образы не были равномерно распределены по всем узлам диаграммы, кроме того, для данной диаграммы характерно большое число узлов, к которым не был отнесён ни один образ.

На всех диаграммах, приведённых на рис. 1, есть нейроны, которые ни разу не среагировали на данные из обучающей выборки, в то же время, другие нейроны среагировали несколько раз, что говорит о том, что в исходном обучающем множестве есть очень схожие образы. В табл. 1 приведены результаты реагирования нейронов на обучающую выборку для карт разного размера.

Таблица 1

Размер карты	Число среагировавших нейронов	Число не среагировавших нейронов
10 × 10	58	42
5 × 5	21	4
3 × 3	8	1

По результатам, приведенным в табл. 1, видно, что число не среагировавших нейронов с уменьшением размеров карты значительно уменьшается.

В случаях, когда число классов заранее известно, число нейронов в сети Кохонена может соответствовать числу классов.

Узлы диаграммы Хинтона также можно представить в виде кластеров [6], тогда образы, попавшие в один узел, будут соответствовать одному классу. На рис. 2 приведена диаграмма Хинтона, представляющая собой 4 узла, т.е. отображающая 4 класса. Оттенки узлов соответствуют степени удаленности их друг от друга, а также сгущению точек, т.е. на данной диаграмме самый удалённый узел карты также соответствует и наименьшему числу образов, попавших в него.

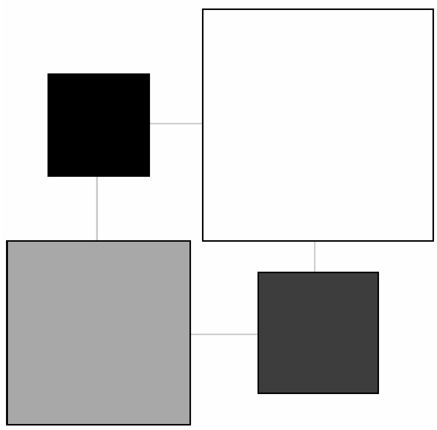


Рисунок 2 – Диаграмма Хинтона для карты Кохонена с 4-мя нейронами

В табл. 2 приведено число попаданий в каждый узел для диаграммы, приведённой на рис. 2.

В табл. 3 приведены наиболее весомые усреднённые значения по результатам сегментации нейронной сетью Кохонена информации о заемщиках в банках.

Таблица 2

Номер класса	Цвет	Число отнесённых образов
1	Тёмно-серый	19
2	Светло-серый	26
3	Белый	38
4	Чёрный	17

По рис. 2 и табл. 2 и 3 можно сделать следующие выводы: самым небольшим по числу принятых образов является четвёртый сегмент, он удалён от трёх других кластеров и точки, находящиеся в нём, – разрежены, кроме того, в нём находится наименьшее количество образов. По результатам табл. 3 видна аналогичная ситуация: это лица в возрасте свыше 60, что вполне соответствует небольшому числу таких заемщиков, у которых есть множество расхождений в характеристиках и эти характеристики почти не пересекаются с характеристиками других классов. Наиболее приближен к нему первый – тёмно-серый сегмент – лица в возрасте до 22 лет. Приближение данных сегментов обусловлено значительным отдалением от других сегментов, также у них есть схожие характеристики, отличные от двух других сегментов, такие как: небольшой среднемесячный доход и расход, незанятость. Данные сегменты не очень схожи друг с другом, их приближение обусловлено значительным отдалением от других сегментов. К первому сегменту менее всего приближен третий сегмент, характеристики этих сегментов значительно отличаются.

Самые крупные по числу отнесённых к ним образов второй – светло-серый, и третий – белый, сегменты, расположенные наиболее близко друг к другу, это лица в возрасте от 22 до 40 лет и от 40 до 60 соответственно. Также по результатам, приведенным в табл. 3, эти сегменты содержат образы с наиболее схожими характеристиками, такими как семейное положение, цели кредита, среднемесячный доход и расход и другие. Расхождения в характеристиках третьего сегмента меньше, чем в трёх других сегментах.

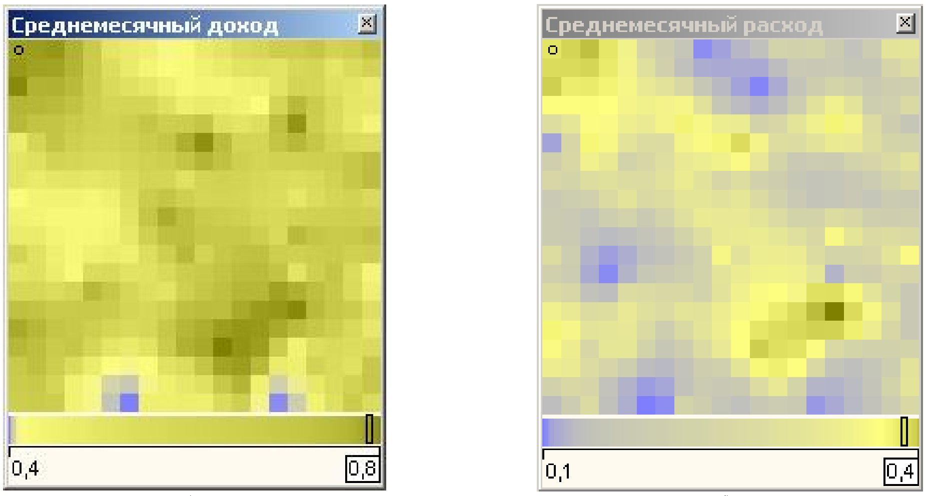
Таблица 3

Характеристики	Усреднённые значения свойств сегментов			
	Сегмент 1	Сегмент 2	Сегмент 3	Сегмент 4
Возраст	до 22 лет	от 22 до 40	от 40 до 60	свыше 60
Семейное положение	холост (не замужем)	женат (замужем)	женат (замужем)	разведён(а), вдовец(а), женат (замужем)
Занятость	нет	да	да	нет
Квартира	нет	да	да	да
Машина	нет	нет	да	да, нет
Среднемесячный доход	до 500	от 1500	от 2000	до 500
Среднемесячный расход	до 400	от 1000	от 1000	до 400
Цель кредита	Образование, потребительские товары	Потребительские товары, транспорт, недвижимость	Транспорт, недвижимость	Потребительские товары, иное

Как показали результаты проведённых исследований, диаграмма Хинтона хорошо отображает разделение исходного множества на известное число сегментов, в данной ситуации – 4, что подтверждают результаты, приведенные в табл. 3.

При анализе также весьма важной является карта входов, т.е. оценка весов нейронов. Для каждого входа нейрона рисуется своя карта, которая раскрашивается в соответствии со значением соответствующего веса нейрона. У нейронной сети, обучаемой с учителем, веса нейронов не имеют физического смысла и не используются при анализе, но при обучении без учителя веса нейронов подстраиваются под значения входных переменных и отражают их внутреннюю структуру. Обычно анализируют одновременно несколько карт входов.

На рис. 3 приведены карты входов размером 20 × 20, на которых образы равномерно распределены по узлам сетки, для таких входов, как среднемесячный доход и расход заемщиков.

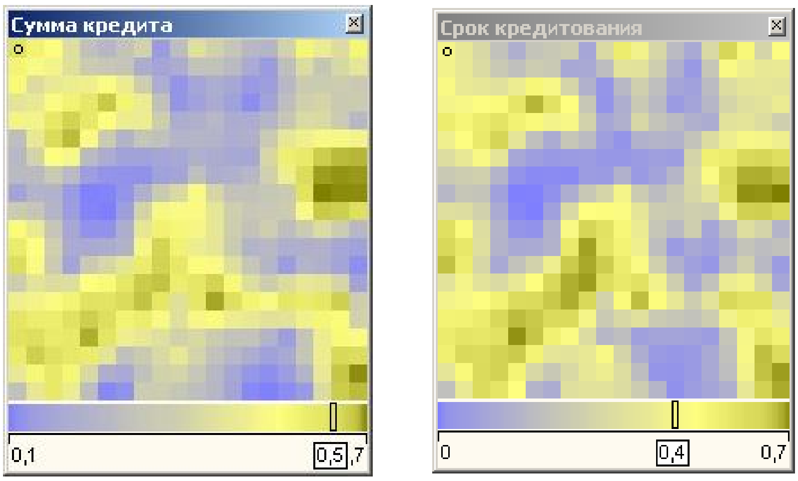


а) б)

Рисунок 3 – Карты входов: а) среднемесячный доход; б) среднемесячный расход

Карты входов (рис. 3) имеют очень схожий вид. Схожесть начертания карт определяет и схожесть данных входов. По данным картам можно сделать вывод, что определённые доходы влекут соответствующие расходы, кроме случаев, в которых одни и те же нейроны на разных картах окрашены в разный цвет.

Иная ситуация с такими входными данными, как сумма кредита и срок, на который он выдан, карты для данных входов приведены на рис. 4.



а) б)

Рисунок 4 – Карты входов: а) сумма кредита; б) срок

По данным компонентам (рис. 4) можно сделать вывод, что для временного периода, по данным за который построены карты, срок кредитования в большинстве случаев непосредственно зависел от выданной заемщику суммы.

Карты Кохонена как аппарат интеллектуального анализа могут применяться в банковских системах для кластеризации, классификации и когда необходима визуализация более чем двумерного входного пространства.

Выводы

Архитектура, основанная на плагинах, позволяет создавать приложения с поддержкой динамического добавления функциональности, что позволит пользователю расширять возможности этого приложения. Система для поддержки банковского кредитования, основанная на плагинах, позволит практически неограниченно наращивать функциональность.

Проведён эксперимент по построению и визуализации карт Кохонена для многомерного набора данных из области банковского кредитования. Данные о клиентах были разделены сетью на 4 группы и отображены посредством диаграммы Хинтона.

Полученные результаты свидетельствуют об эффективности применения аппарата нейронных сетей Кохонена для визуализации данных различной размерности, с различными структурами.

Перспективы и дальнейшие исследования направлены на применение средств визуализации к другим нейронным сетям, решающим аналогичные задачи.

Литература

1. Определение плагина и архитектуры, основанной на плагинах [Электронный ресурс]. – Режим доступа : http://en.wikipedia.org/wiki/Plugin.
2. Калан Р. Основные концепции нейронных сетей / Калан Р.; пер. с англ. – М. : Издательский дом «Вильямс», 2003. – 288 с.
3. Хайкин С. Нейронные сети: полный курс / Хайкин С. – 2-е изд. – М. : ООО «Вильямс», 2006. – 1104 с.
4. Зиновьев А.Ю. Визуализация многомерных данных / Зиновьев А.Ю. – Красноярск : Изд-во КГТУ, 2000. – 168 с.
5. Kohonen T. Self-organizing maps / Teuvo Kohonen. – 3 ed. – Berlin : Springer-Verlag Berlin Heidelberg, 2001. – 500 p.
6. Бодянский Е.В. Искусственные нейронные сети: учебное пособие / Е.В. Бодянский, О.Г. Руденко. – Харьков : ООО «Компания СМИТ», 2005. – 408 с.

А.С. Коряк

Інформаційно-аналітична система підтримки банківського кредитування, що заснована на плагінах

У статті наведений опис інформаційно-аналітичної системи підтримки банківського кредитування, що заснована на плагінах. Для системи, що описана, розроблені плагіни, які мають алгоритми навчання, роботи та візуалізації нейронних мереж, що призначені для створення моделей кластеризації, класифікації та прогнозування.

Статья поступила в редакцию 16.07.2009.